

HUMAN DISEASE DETECTION

G.SAI SRUJAN, B.DEVADAS, V.SAKETH

Ms. NARRA SHIVANI (Guide)

Sreenidhi Institute of Science and Technology

Hyderabad

ABSTRACT:

"Human Disease Detection Using Machine Learning" predicts a person's disease based on the data or symptoms that person enters into the system and then delivers correct findings based on that data. If the user merely wants to know the type of ailment the patient has experienced and the patient is not too seriously ill. It is a system that gives users advice on how to keep their health systems in good shape and offers a technique to identify diseases using this prediction. Today, the health sector plays a significant role in treating patients' diseases, so this system is also helpful for the sector to inform the user and is beneficial to the user if he or she chooses not to visit a hospital or other clinic. By entering symptoms and other pertinent information, the user can learn what disease they are experiencing, and the sector can

benefit from this system by simply asking the user.

INTRODUCTION

The entire process of this Human Disease Detection Using Machine Learning was carried out with the use of machine learning, the Python programming language, the Tkinter interface, and the dataset made previously available by the hospitals, which will be used to predict the disease. Today's doctors use a variety of scientific technology and methodologies for both the detection and diagnosis of numerous fatal diseases as well as ordinary illnesses. Correct and thorough diagnosis is always responsible for a successful course of treatment. Ailment prediction systems that use machine learning algorithms help in such situations to generate reliable findings because doctors can occasionally make mistakes while diagnosing a patient's

disease. The project disease prediction using machine learning is developed to combat general disease in earlier stages as we all know in competitive environment of economic development the mankind has been so involved that he/she is not concerned about health according to research there are 40% peoples who ignore about general disease which leads to harmful disease later. The biggest cause of ignorance is people's reluctance to see a doctor and their excessive time commitments, which leaves them with no time to schedule an appointment and visit a physician, ultimately leading to fatal illness. According to study, 25% of deaths in India are attributed to early ignorance, and 70% of people suffer from general sickness.

PROBLEM DEFINITION

In the health industry today, there are many issues with machines or devices that will produce incorrect or unacceptable results. In order to avoid these results and obtain the correct and desired results, we are developing a programme or project that will produce accurate predictions based on data that has been supplied by the user as well as data that is already present in that machine. The health sector is a very large industry with a lot of unfinished business and is still underdeveloped in terms of knowledge and

information. So, we completed this project that will assist those in need with the aid of all those algorithms, approaches, and methodologies. The issue is that many people go to hospitals or clinics to find out how their health is and how much they have improved over the past few days, but they have to travel to get those answers. Additionally, patients may or may not receive the results depending on a number of variables, including the fact that the doctor may be out sick or experiencing another issue, which may have prevented him from visiting the hospital, among other possibilities. In order to avoid these complications and reasons, we are now implementing new procedures. And occasionally, if a person has been exhibiting a few symptoms but is unsure of the illness they are dealing with, this can result in a variety of ailments in the future. Therefore, this disease prediction will be very helpful to a variety of people, ranging from children to teenagers to adults and also senior citizens, in order to avoid that and learn about the condition in early stages of the symptoms.

OBJECTIVE OF THE PROJECT

A system that would make it simple for end users to forecast diseases without visiting a doctor or physician for a diagnostic needs

to be researched and developed. Machine learning will improve the accuracy of predictions.

PURPOSE

The goal of this research, "Human Disease Detection," is to accurately forecast the patient's disease using all of their general information as well as their symptoms. Using this data, we will compare with patient datasets from the past and make predictions about the disease the patient will eventually experience. In general, this prediction system can also be very helpful in the health industry. If this forecast is made at the early stages of the disease with the aid of this project and all other essential measures, the condition can be cured. The job of the doctors can be decreased and they will be able to accurately forecast the patient's sickness if the health sector embraces this idea. The main goal of this disease prediction is to offer predictions for a variety of often occurring illnesses that, if left untreated or sometimes even neglected, can become fatal illnesses and cause significant problems for both the sufferer and their loved ones. Based on the symptoms, this method will identify the most probable disease. The health sector is a very large industry with a lot of unfinished business and is still underdeveloped in

terms of knowledge and information. So, we completed this project that will assist those in need with the aid of all those algorithms, approaches, and methodologies.

SCOPE

The project's goal is to integrate clinical decision assistance into computerised patient records in order to prevent medical errors, increase patient safety, reduce unneeded practise variation, and enhance patient outcomes. As tools for data modelling and analysis, this concept shows promise.

OVERALL DESCRIPTION The study of computer algorithms that get better on their own with practise is known as machine learning (ML). It is considered to be a component of artificial intelligence. Without being expressly taught to do so, machine learning algorithms create a model using sample data, also referred to as "training data", in order to make predictions or judgements. In a wide range of applications, including email filtering and computer vision, when the development of traditional algorithms to carry out the required tasks is challenging or impractical, machine learning techniques are used.

IMPLEMENTATION

A module is a group of source files and build options that enables you to segment the functionality of your project into discrete functional pieces. Our project may have a number of modules, and one module may depend on another module. It is possible to individually build, test, and debug each module. In project open, a module is a high-level description of a functional area made up of a set of processes that describe the functionality of the module and a set of packages that actually implement it. In our project, we have five components, including:

- User Module
- Pre-processing Module
- Training Module
- Detection Module
- Testing Module

USER MODULE

In the user module, the user provides the system with input in the form of symptoms, and the system preprocesses those symptoms to produce an output in the form of disease.

PRE-PROCESSING MODULE

The symptoms are preprocessed in this module in order to extract the useful information from the provided data.

TRAINING MODULE

The model needs to be trained after it has been generated. We have a predetermined dataset with the diseases and symptoms in it. This dataset is used to train the model.

TESTING MODULE

The model must be tested after it is created. We have a predetermined dataset with the diseases and symptoms in it. The model is evaluated using this dataset.

DETECTION MODULE

In this module, the model forecasts a user's illness. A dialogue box will be displayed if the user enters one symptom or if they don't submit any symptoms at all. If the user provides a minimum of two symptoms or a maximum of five symptoms, the disease is returned as an output.

IMPORT LIBRARIES

We're using a number of libraries in this project, including Tkinter, NumPy, Pandas, and Sklearn.

Tkinter: The Python binding for the Tk GUI toolkit is called Tkinter. It serves as the de facto default GUI for Python and the standard Python interface to the Tk GUI toolkit. It comes with normal Python installations on Linux, Windows, and Mac OS X.

Sklearn: The most effective and reliable Python machine learning library is Scikit-learn (Sklearn). Through a Python consistency interface, it offers a variety of effective tools for statistical modelling and machine learning, including classification, regression, clustering, and dimensionality reduction.

ALGORITHMS

Our research makes use of two algorithms. As follows:

- Forest Random
- K-Nearest Neighbor

RANDOM FOREST

A large number of decision trees are built during the training phase of the random forests or random decision forests ensemble learning approach, which is used for classification, regression, and other tasks. The class that the majority of the trees chose is the output of the random forest for classification problems. The mean or average prediction of each individual tree is returned for regression tasks. The tendency of decision trees to overfit their training sets is corrected by random decision forests. Although they frequently outperform decision trees, gradient boosted trees are more accurate than random forests.

However, their effectiveness may be impacted by data peculiarities. The Random Forest Algorithm's ability to handle data sets with both continuous variables, as in regression, and categorical variables, as in classification, is one of its most crucial qualities. In terms of classification issues, it delivers superior outcomes.

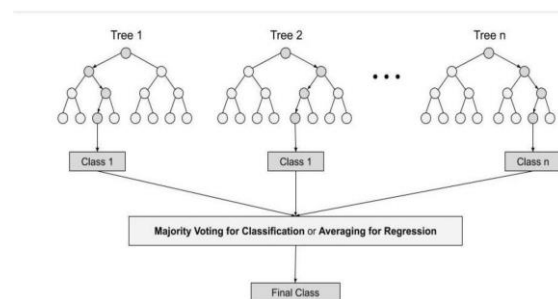


Fig1: Random Forest

KNN

KNN is an easy-to-use, supervised machine learning (ML) technique that is frequently used in missing value imputation. It can be utilised for classification or regression problems. It is predicated on the notion that the observations most "similar" to a given data point are those that are closest to it in the data set, allowing us to categorise unanticipated points based on the values of the existing points that are closest to them. The user can choose K to specify how many

neighbouring observations will be used in the algorithm.

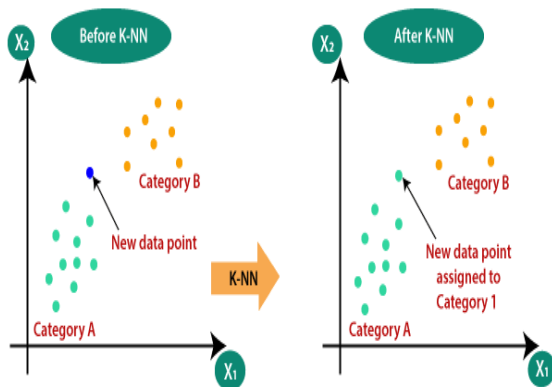


Fig2: KNN

RESULTS

It forecasts the disease by providing the patient's name and at least two symptoms.

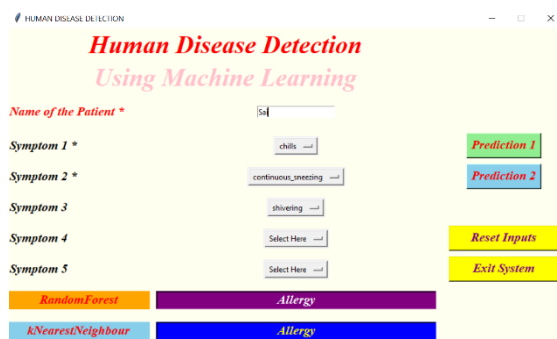


Fig3: Result

In the absence of a patient name, it instructs to enter the patient name.

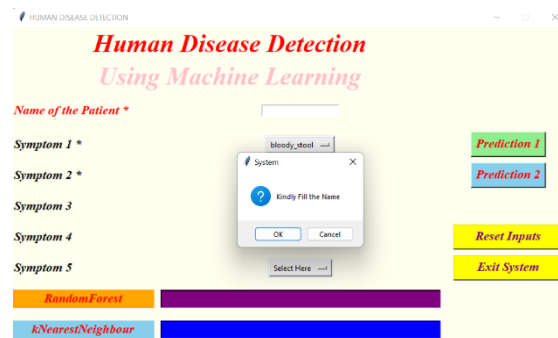


Fig4: Result when no patient name is provided

It instructs to enter at least two symptoms if a minimum of two symptoms are not provided.

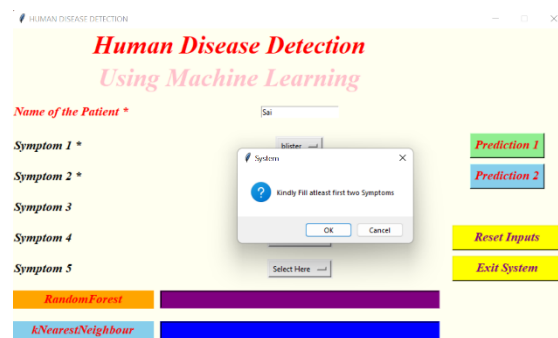


Fig5: Result when less than two symptoms are provided

MERITS

In terms of merits, our project's increased accuracy makes it more dependable for forecasting. In forecasting a patient's disease, it helps the doctor's task a little.

SOFTWARE AND HARDWARE REQUIREMENTS

For this project, we used Windows 10, a 2.40 GHz Intel Core i5-1135G7 processor, with 16.0 GB of RAM.

CONCLUSION AND FUTURE SCOPE

This undertaking Everyday life benefits greatly from disease prediction using machine learning, but it is particularly important for the healthcare industry because it is they who use these systems daily to predict the diseases of the patients based on their general information and their symptoms that they have been through. These days, the health sector plays a significant role in treating patients' illnesses, thus this is a way for the sector to inform users and is also helpful to users who don't want to visit hospitals or other clinics. Therefore, the user can learn what disease he or she has simply by entering the symptoms and any other pertinent information. The health industry can also benefit from this system by simply asking the user to enter their symptoms, which the system will then do in a matter of seconds, in order to identify the exact and, to some extent, accurate diseases. The job of the doctors can be decreased and they will be able to accurately forecast the patient's sickness if the health sector

embraces this idea. The goal of the disease prediction is to offer forecasts for a wide range of often recurring illnesses that, if left untreated or occasionally overlooked, can progress to fatal conditions and cause significant problems for both the patient and their loved ones.

ACKNOWLEDGMENTS

We would especially want to thank Ms.Narra Shivani, our mentor, for providing us with the chance to work on this amazing project on this subject. She also assisted us in conducting extensive research, which allowed us to learn a lot of brand-new information. We are very appreciative to them.

Second, we also want to express our gratitude to my friends, who greatly contributed to the timely completion of this job.